



HACETTEPE ÜNİVERSİTESİ
İST366-REGRESYON ÇÖZÜMLEMESİ ÖDEVİ

SEDANUR GÜLTEN-21935965

İÇİNDEKİLER TABLOSU

Tanımlayıcı istatistikler	3
Normallik incelemesi.....	4
Doğrusallık.....	6
Regresyon Modeli	7
Aykırı Değer İncelemesi	8
Aykırı Değerlerden Sonraki Analiz Kısmı	11
İkinci Aykırı Değer İncelemesi.....	14
Model Anlamlılığı, Belirtme Katsayısı, Katsayı Yorumları ,%99 Güven Düzeyi.....	17
Değişen Varyanslılık	18
Öz ilişki ve Çoklu Bağlantı Sorunu.....	19
Uyum kestirimi ve Ön kestirim.....	21
Değişken Seçimi Yöntemi	22
Ridge Regresyon ve İz Grafiği.....	25

Verileri Okutma

```
rm(list=ls())
regresyon_verii<-read_xlsx("C:/Users/sedad/Desktop/regresyon_veri.xlsx")
view(regresyon_verii)
names(regresyon_verii)<-c("y","x1","x2","x3","x4")
names(regresyon_verii)
x4<-as.factor(x4)
class(regresyon_verii$x4)
attach(regresyon_verii)
```

Tanımlayıcı İstatistikler

```
install.packages("psych")
library(psych)
ozet<-describe(regresyon_verii)
view(ozet)
data.frame(ozet)
```

- Özet İstatistikler

	vars	n	mean	sd	median	trimmed	mad
"y"	1	150	30.11419172	13.0533974	28.5951652	28.7801073	9.9280891
"x1"	2	150	6.00590576	1.0301814	6.0072272	6.0028232	0.9925248
"x2"	3	150	0.09759413	1.0131760	0.2012332	0.1052147	1.0348731
"x3"	4	150	3.11236257	0.9756782	3.1873385	3.1184228	1.1542526
"x4"	5	150	2.00000000	0.8192319	2.0000000	2.0000000	1.4826000
	min	max	range	skew	kurtosis	se	
"y"	9.5100645	92.197618	82.687553	1.62848245	4.6502484	1.06580544	
"x1"	3.8969804	8.465775	4.568794	0.07050762	-0.3681134	0.08411396	
"x2"	-2.3727331	2.378632	4.751365	-0.13470066	-0.6474996	0.08272548	
"x3"	0.4070959	5.404671	4.997575	-0.08930516	-0.6177941	0.07966379	
"x4"	1.0000000	3.000000	2.000000	0.00000000	-1.5199333	0.06689001	

- Betimsel İstatistikler

```
> summary(regresyon_veri)
      y          x1          x2          x3
Min.   : 9.51   Min.   :3.897   Min.   : -2.37273   Min.   :0.4071
1st Qu.:21.87   1st Qu.:5.329   1st Qu.: -0.66178   1st Qu.:2.3296
Median :28.60   Median :6.007   Median : 0.20123   Median :3.1873
Mean   :30.11   Mean   :6.006   Mean   : 0.09759   Mean   :3.1124
3rd Qu.:35.17   3rd Qu.:6.632   3rd Qu.: 0.84012   3rd Qu.:3.8693
Max.   :92.20   Max.   :8.466   Max.   : 2.37863   Max.   :5.4047

      x4
Min.   :1
1st Qu.:1
Median :2
Mean   :2
3rd Qu.:3
Max.   :3
```

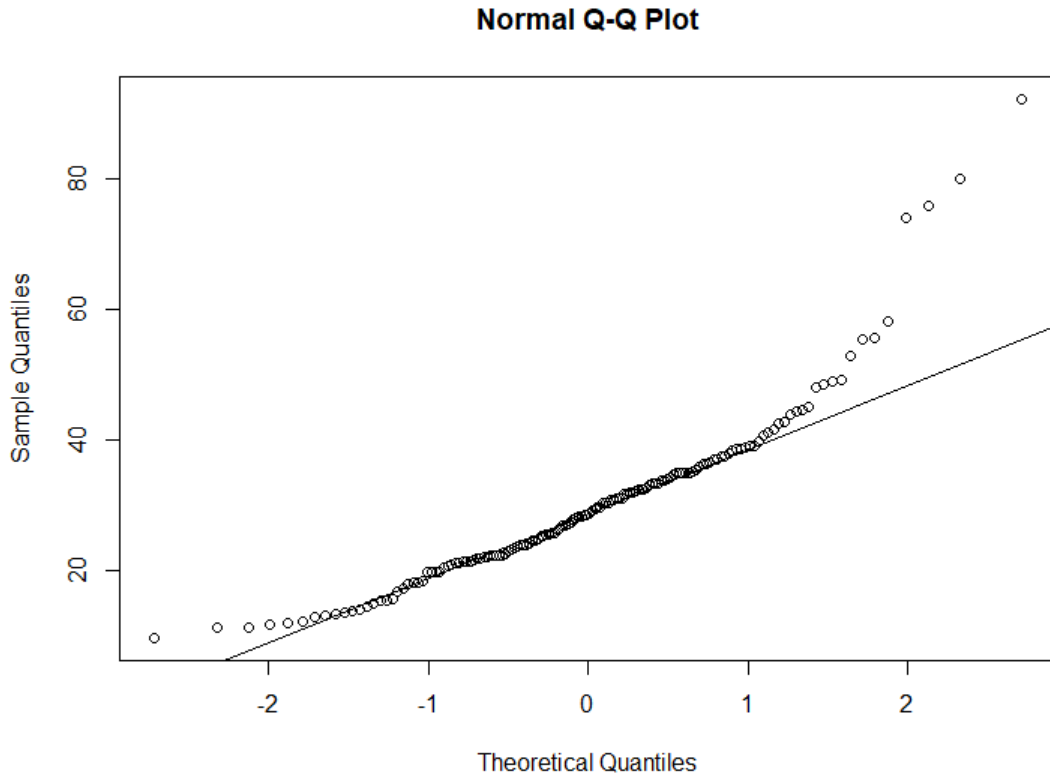
Betimsel İstatistik Tablosu:

"y"	"x1"	"x2"	"x3"	"x4"
Min. :9.51	Min. :3.897	Min. :-2.37273	Min. :0.4071	Min. :1
1st Qu.:21.87	1st Qu.:5.329	1st Qu.: -0.66178	1st Qu.:2.3296	1st Qu.:1
Median :28.60	Median :6.007	Median : 0.20123	Median :3.1873	Median :2
Mean :30.11	Mean :6.006	Mean : 0.09759	Mean :3.1124	Mean :2
3rd Qu.:35.17	3rd Qu.:6.632	3rd Qu.: 0.84012	3rd Qu.:3.8693	3rd Qu.:3
Max. :92.20	Max. :8.466	Max. : 2.37863	Max. :5.4047	Max. :3

QQNorm Grafiği

```
install.packages("graphics")
library(graphics)
qqnorm(y)
qqline(y)
install.packages("lmtest")
library(lmtest)
install.packages("nortest")
library(nortest)
lillie.test(y)
```

Y bağımlı değişkenimizin normal dağılıma uyumu:



Örneklem büyüklüğü 30 ve daha fazla olması durumunda kolmogrov simirnov testi önerilmektedir

H_0 : Verilerin dağılışı ile normal dağılım arasında fark yoktur.

H_s : Verilerin dağılışı ile normal dağılım arasında fark vardır.

P değerimiz = 0.0007541 $\Rightarrow p < \alpha = 0.05$ olduğu için veriler normal dağılmıyor. Dönüşüm uygulanmalı.

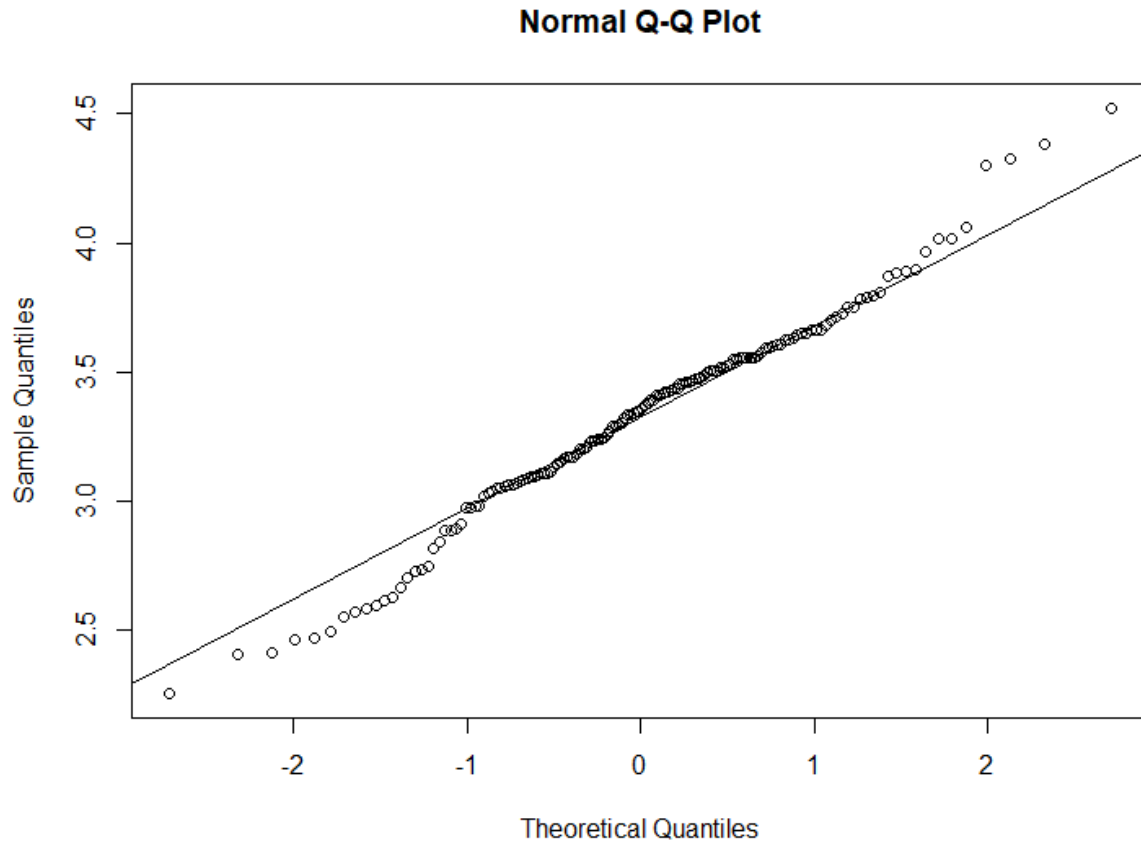
```
> library(nortest)
> lillie.test(y)

      Lilliefors (Kolmogorov-Smirnov) normality test

data:  y
D = 0.10091, p-value = 0.0007541
```

Dönüşüm sonrası:

```
lmy<-log(y)
qqnorm(lmy)
qqline(lmy)
lillie.test(lmy)
```



Lilliefors (Kolmogorov-Smirnov) normality test

data: lny

D = 0.054814, p-value = 0.3298

Lilliefors (Kolmogorov-Smirnov) normality test

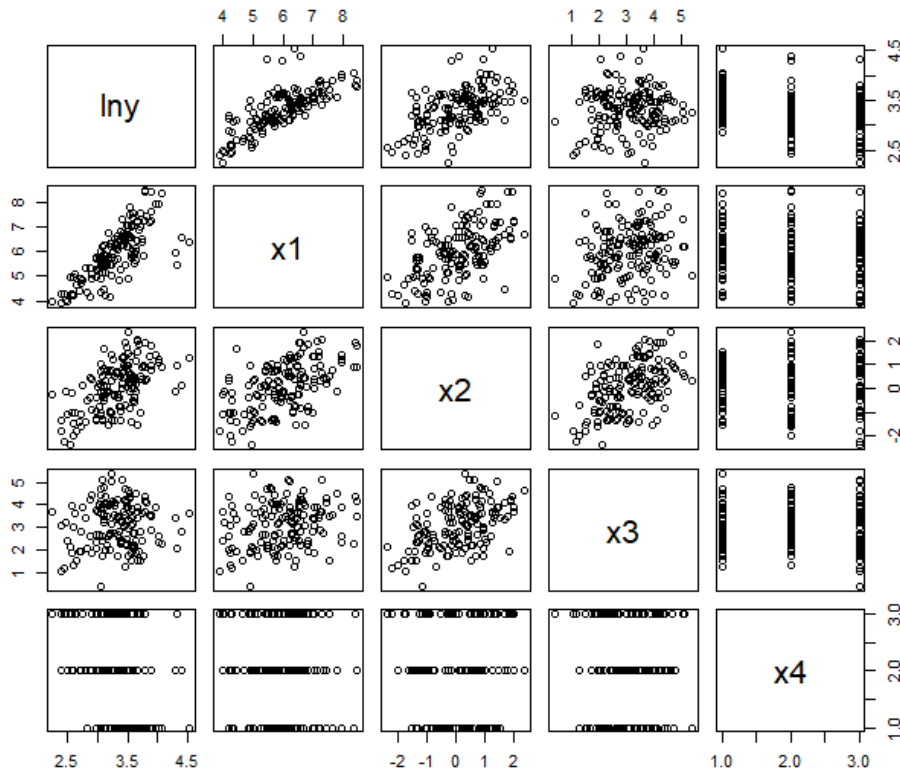
data: lny

D = 0.054814, p-value = 0.3298

P değeri=0.3298>0.05 olduğundan bağımlı y değişkenimiz normal dağılmaktadır.

Şimdi doğrusallık incelemesi yapılmalıdır.

```
regresyon_veriyeni<-cbind(lny,x1,x2,x3,x4)
view(regresyon_veriyeni)
pairs(regresyon_veriyeni)
```



Bağımlı değişken lny için doğrusallık incelendiğinde bağımsız değişkenlerden x1,x2,x3 değişkenleri ile doğrusal ilişki olduğu görülür ,doğrusal model oluşturulabilir.Ancak bağımsız değişkenler birbirleriyle ilişkili durumdadır.x4 değişkenimiz nitel değişken olduğundan doğrusallık incelemesinde kullanılmaz.

Regresyon modeli:

```
sonuc<-lm(lny~x1+x2+x3+x4)
summary(sonuc)
```

```
Call:
lm(formula = lny ~ x1 + x2 + x3 + x4)

Residuals:
    Min       1Q   Median       3Q      Max
-0.34142 -0.08488  0.00290  0.03853  1.11327

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.39453    0.13358   17.925 < 2e-16 ***
x1           0.23467    0.01912   12.276 < 2e-16 ***
x2           0.15066    0.02142    7.033 7.51e-11 ***
x3          -0.09845    0.01937   -5.081 1.14e-06 ***
x42          -0.24225    0.04074   -5.946 1.99e-08 ***
x43          -0.33070    0.04063   -8.139 1.70e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2017 on 144 degrees of freedom
Multiple R-squared:  0.7675,    Adjusted R-squared:  0.7594
F-statistic: 95.06 on 5 and 144 DF,  p-value: < 2.2e-16

> |
```

$$y=2.39453+0.23467x_1+0.15066x_2-0.09845x_3-0.24225x_{42}-0.33070x_{43}\pm 0.2017$$

$$(0.13358) (0.01912) (0.02142) (0.01937) (0.04074) (0.04063)$$

Bir sonraki sayfada regresyon modelimiz için aykırı değer incelemesi yapacağız.

Aykırı Değer İncelemesi

```
sonuc<-lm(lm~x1+x2+x3+x4)
summary(sonuc)
predict(sonuc)
inf<-ls.diag(sonuc)
inf
```

```
> inf<-ls.diag(sonuc)
> inf
$std.dev
[1] 0.2017353
```

h_{ii} (Gözlem Uzaklığı: Uç Değerler)

```
$hat
[1] 0.05301170 0.02251243 0.02575581 0.06184709 0.03468202 0.03653451 0.03988330
[8] 0.04206980 0.05247628 0.03120945 0.02617344 0.02478057 0.08100287 0.02095393
[15] 0.02324266 0.06191806 0.05889225 0.05448290 0.06382023 0.03078578 0.03978770
[22] 0.07230664 0.05094198 0.04873860 0.04025504 0.03188322 0.03413250 0.03151140
[29] 0.04684383 0.02681498 0.04939981 0.03171436 0.05171476 0.04256994 0.03614158
[36] 0.02723877 0.02447346 0.03551171 0.02548486 0.02605789 0.02326227 0.02168600
[43] 0.04550496 0.04728718 0.03274790 0.03538011 0.03243196 0.02281448 0.03324564
[50] 0.03053170 0.03270563 0.07966119 0.03226314 0.03280544 0.02211091 0.05713053
[57] 0.03341061 0.03590603 0.02378398 0.09294462 0.04079576 0.04430756 0.02125596
[64] 0.03199082 0.03619187 0.06643044 0.03447035 0.06723773 0.03115042 0.02422555
[71] 0.03958788 0.04318467 0.04730588 0.06066394 0.03410215 0.02994908 0.03699009
[78] 0.04678812 0.03515143 0.02531448 0.04928645 0.04110220 0.05163193 0.02646463
[85] 0.03093340 0.02281777 0.04025761 0.03990405 0.03767204 0.07265734 0.02527690
[92] 0.02841733 0.03597925 0.03867542 0.06092376 0.02609583 0.03967111 0.02642000
[99] 0.04216771 0.03455421 0.04438971 0.03856409 0.03115019 0.06132962 0.02662232
[106] 0.03864365 0.07201713 0.06627787 0.03289933 0.03619926 0.05464382 0.02450238
[113] 0.04090736 0.06603092 0.04323046 0.03004788 0.04881499 0.06555780 0.07125608
[120] 0.02541915 0.02495784 0.05528039 0.03530762 0.03148744 0.04830804 0.03042568
[127] 0.03274843 0.04147414 0.04189162 0.02889872 0.02148986 0.02965177 0.02196276
[134] 0.06169353 0.04968830 0.03490763 0.04319313 0.03867427 0.06370080 0.03674819
[141] 0.02762010 0.02232383 0.05032184 0.04847237 0.02183496 0.02807852 0.03812963
[148] 0.03732187 0.04398388 0.03574324
```

Hii>(2(k+1))/n olduğu durumda uç değer olur.(k=4,n=150) hii=(2(5))/150=0,067 değeridir.

Hii 0,067 değerinden yüksek değerler olarak 13,22,52,60,68,90,107 ve 119. Gözlemler> 0,067 olduğundan uç değerlerdir.

Std.res(Aykırı Değerler=r_j)

```
$std.res
[1] 0.5459018055 0.4785769626 0.5481655975 0.7724821500 -0.3054107633 0.4310183979
[7] 0.0544387695 0.8191493642 -0.8605631049 0.7977926852 0.5872100398 0.5405201480
[13] 0.7629224784 0.5027444307 0.4585406893 0.9478522109 0.0571740595 0.6855233995
[19] -1.0172085555 4.0280204427 0.3200780167 0.7434716081 0.5950845768 0.5304842681
[25] 0.9255107725 0.0979458994 0.9312891437 -0.3352128047 0.9063730476 0.1522026479
[31] -0.8719552942 -0.3242436419 0.7806392350 -0.6635843694 0.4028030472 0.6009853804
[37] 0.1504505793 -1.1444941127 -1.2417580829 -1.0264709247 -1.1697038328 -1.1182477840
[43] -1.0058727737 -1.4798132763 -1.2970760589 -1.5908017104 -1.2058246917 -1.0222321018
[49] -1.1567453336 -1.2601003260 0.2635763231 0.3573887469 -0.0669039636 0.2015358668
[55] 0.0561623981 -0.5240143284 0.1933487738 -0.2123711705 0.0006224013 -0.1194825824
[61] -0.2514998462 -0.9788563670 5.5730794718 0.0150957327 0.0042494285 -0.9178450615
[67] 0.1946683811 -0.4379037683 0.0138136389 0.0477834693 -0.0839017093 -0.0612109773
[73] -0.4328006220 -0.2644148827 0.0949977614 0.0425075077 0.2782293814 -0.8963836339
[79] -0.1469051626 4.7570111300 0.0760220320 -1.2097368430 -0.0638110455 -0.0770618179
[85] -0.4052615404 0.0496213345 -0.4134844709 -0.2098521395 0.1729311225 0.1389857660
[91] -0.2958383060 -0.3851988755 -1.0653314677 -0.4230711959 -0.6138361854 -0.4383173152
[97] -0.7141004598 -0.4324936982 -0.0843367467 -0.4231269294 -0.6734329562 0.1432468146
[103] 0.1460584026 0.1797458219 0.1734687334 0.1896352278 0.2874926164 -1.0415927861
[109] 0.0504017360 0.1352394616 0.0522958001 0.2588268882 -0.0777104847 -0.5971299333
[115] 0.1278408446 -0.0825534678 0.4000583299 -0.8312621488 -0.3965407515 0.0562552744
[121] 0.0602669574 -1.7412292277 -0.0083739401 0.0791560436 5.2754706234 -0.0913660721
[127] 0.1661469608 -0.0735159999 -0.6096096705 0.0140995602 0.1239186165 0.0735402173
[133] 0.1474924858 -0.5667595071 0.0413780567 -0.0378525531 -0.3212726512 0.3404818938
[139] -0.2078763046 0.0820962925 0.2181935770 0.1460573890 -0.3788393135 -1.0916138734
[145] 0.1769457402 0.1463633822 0.0476321017 -0.4543818969 -0.0525010375 -0.0583820409
```


Standartlaştırılmış hataların (-2,+2) aralığında olması istenir. Bu aralığa girmeyen gözlemler:20,63,80 ve 125.gözlemlerdir.

Stud.res (Student Tipi Artıklar= r_i)

```
$stud.res
[1] 0.5445667947 0.4772920661 0.5468297629 0.7713952180 -0.3044470803 0.4297965323
[7] 0.0542499750 0.8182086951 -0.8597835471 0.7967805619 0.5858694320 0.5391873293
[13] 0.7618100111 0.5014360082 0.4572797276 0.9475157496 0.0569758393 0.6842564070
[19] -1.0173320417 4.2612464526 0.3190782241 0.7423116791 0.5937452314 0.5291564075
[25] 0.9250469751 0.0976084684 0.9308573374 -0.3341772534 0.9058079241 0.1516854473
[31] -0.8712254339 -0.3232338521 0.7795752619 -0.6622896412 0.4016283176 0.5996474857
[37] 0.1499390562 -1.1457361340 -1.2441178945 -1.0266635052 -1.1712126675 -1.1192284311
[43] -1.0059142071 -1.4860084389 -1.3001820501 -1.5993845228 -1.2077434587 -1.0223928320
[49] -1.1581150671 -1.2626983863 0.2627229158 0.3563037060 -0.0666722897 0.2008632002
[55] 0.0559676630 -0.5226902565 0.1927012694 -0.2116656359 0.0006202364 -0.1190728921
[61] -0.2506801233 -0.9787131971 6.2781912449 0.0150432376 0.0042346481 -0.9173398283
[67] 0.1940168026 -0.4366714669 0.0137656004 0.0476176429 -0.0836119201 -0.0609988625
[73] -0.4315760141 -0.2635591656 0.0946703000 0.0423599208 0.2773361794 -0.8957684072
[79] -0.1464051577 5.1635097129 0.0757591271 -1.2117020006 -0.0635899925 -0.0767953595
[85] -0.4040824314 0.0494491608 -0.4122910865 -0.2091542009 0.1723475182 0.1385116265
[91] -0.2948989305 -0.3840569671 -1.0658344335 -0.4218619041 -0.6125029685 -0.4370844008
[97] -0.7128799859 -0.4312695606 -0.0840454763 -0.4219175475 -0.6721498388 0.1427587351
[103] 0.1455611544 0.1791407147 0.1728834266 0.1889992261 0.2865748939 -1.0419021806
[109] 0.0502268682 0.1347776215 0.0521143960 0.2579866350 -0.0774418104 -0.5957910415
[115] 0.1274034095 -0.0822682712 0.3988885494 -0.8303654804 -0.3953774054 0.0560602193
[121] 0.0600580900 -1.7537331531 -0.0083448153 0.0788824335 5.8530648563 -0.0910509153
[127] 0.1655849287 -0.0732616661 -0.6082746835 0.0140505278 0.1234941786 0.0732858006
[133] 0.1469905703 -0.5654191473 0.0412343777 -0.0377210793 -0.3202699797 0.3394342645
[139] -0.2071843434 0.0818126539 0.2174705929 0.1455601442 -0.3777098789 -1.0923459939
[145] 0.1763494477 0.1458651408 0.0474667982 -0.4531263894 -0.0523189254 -0.0581796609
```

Student tipi artıkların (-3,+3) aralığında olması istenen durumdur. Bu aralığa girmeyen gözlemler:20,63,80 ve 125.gözlemlerdir.

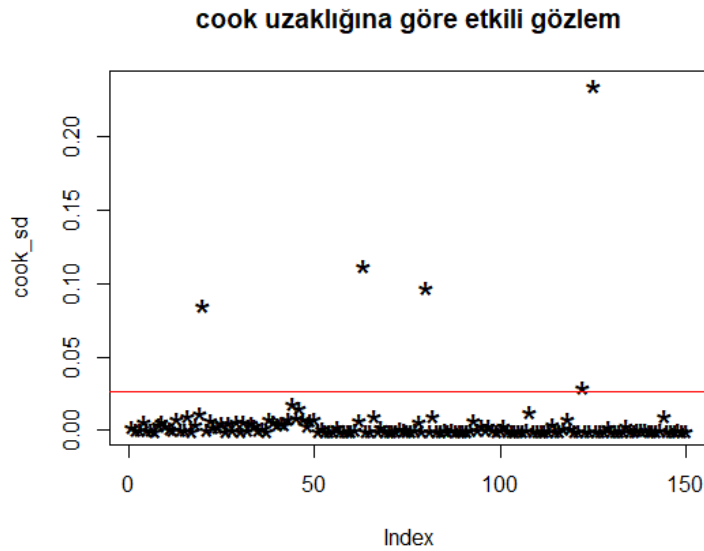
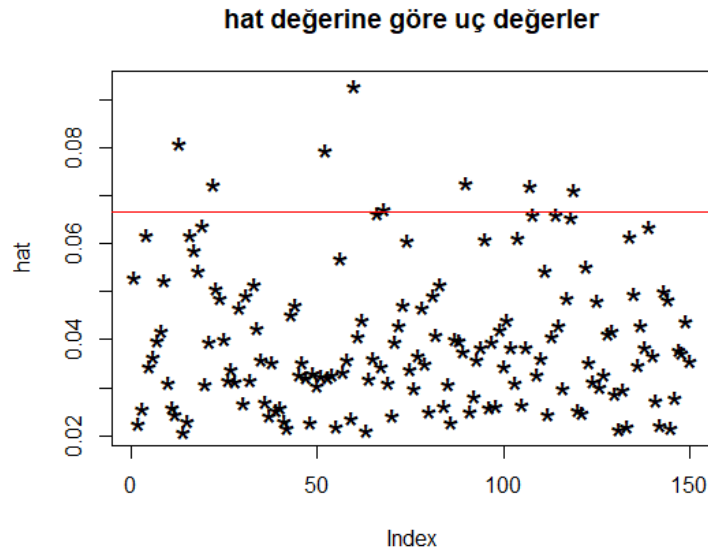
Cook Uzaklığı(D_i)

```
$cooks
[1] 2.780385e-03 8.791510e-04 1.323975e-03 6.556489e-03 5.585363e-04 1.174107e-03
[7] 2.051788e-05 4.911471e-03 6.835765e-03 3.417315e-03 1.544596e-03 1.237318e-03
[13] 8.550585e-03 9.015829e-04 8.338802e-04 9.883405e-03 3.409303e-05 4.513195e-03
[19] 1.175622e-02 8.589394e-02 7.075253e-04 7.180443e-03 3.168029e-03 2.403073e-03
[25] 5.987921e-03 5.265695e-05 5.108206e-03 6.093449e-04 6.729008e-03 1.063837e-04
[31] 6.585133e-03 5.739105e-04 5.538919e-03 3.263150e-03 1.013977e-03 1.685613e-03
[37] 9.464393e-05 8.038047e-03 6.720730e-03 4.698379e-03 5.430935e-03 4.619831e-03
[43] 8.039331e-03 1.811524e-02 9.493435e-03 1.546977e-02 8.122857e-03 4.066130e-03
[49] 7.669073e-03 8.334438e-03 3.914943e-04 1.842594e-03 2.487149e-05 2.296072e-04
[55] 1.188658e-05 2.773012e-03 2.153644e-04 2.799548e-04 1.572998e-09 2.438082e-04
[61] 4.483614e-04 7.403659e-03 1.126237e-01 1.255172e-06 1.130135e-07 9.990977e-03
[67] 2.254860e-04 2.303818e-03 1.022522e-06 9.447749e-06 4.836097e-05 2.818441e-05
[73] 1.550194e-03 7.525410e-04 5.310386e-05 9.297561e-06 4.955750e-04 6.573289e-03
[79] 1.310409e-04 9.795387e-02 4.993506e-05 1.045498e-02 3.694723e-05 2.690552e-05
[85] 8.737628e-04 9.582596e-06 1.195255e-03 3.050548e-04 1.951152e-04 2.522486e-04
[91] 3.782685e-04 7.233063e-04 7.059663e-03 1.200164e-03 4.074174e-03 8.579875e-04
[97] 3.510926e-03 8.459984e-04 5.218829e-05 1.067980e-03 3.511065e-03 1.371770e-04
[103] 1.143158e-04 3.518224e-04 1.371689e-04 2.409243e-04 1.069050e-03 1.283499e-02
[109] 1.440307e-05 1.144901e-04 2.634681e-05 2.804462e-04 4.292881e-05 4.201470e-03
[115] 1.230752e-04 3.518705e-05 1.368938e-03 8.079727e-03 2.010714e-03 1.375683e-05
[121] 1.549498e-05 2.956847e-02 4.277481e-07 3.395072e-05 2.354476e-01 4.365941e-05
[127] 1.557703e-04 3.897499e-05 2.708102e-03 9.859934e-07 5.620695e-05 2.754366e-05
[133] 8.141792e-05 3.519989e-03 1.492028e-05 8.637549e-06 7.765809e-04 7.772995e-04
[139] 4.899919e-04 4.285406e-05 2.253834e-04 8.118382e-05 1.267474e-03 1.011719e-02
[145] 1.164848e-04 1.031470e-04 1.498975e-05 1.334053e-03 2.113551e-05 2.105758e-05
```

$n > 50$ olduğu durumlarda cook uzaklığı D_i değeri $4/n$ den büyük ise etkili gözlem olduğu söylenir. ($n=150$) $\rightarrow 4/150 = 0,0267$ değerinden büyük olan 20,63,80,122 ve 125.gözlemler etkin değerlerdir. Ayrıca 20,63,80 ve 125.gözlemler etkin- aykırı değerlerdir.

Aykırı Değerler İçin Grafikler

```
influence.measures(sonuc)
n<-150
k<-4
cook_sd<-cooks.distance(sonuc)
plot(cook_sd,pch="*",cex=2,main="cook uzaklığına göre etkili gözlem")
abline(h=if(n>50) 4/n else 4/(n-k-1),col="red")
install.packages("zoo")
library(zoo)
hat<-(inf$hat)
plot(hat,pch="*",cex=2,main="hat değerine göre uç değerler")
abline(h=2*(k+1)/n,col="red")
```



Measure influence

```
inf  
influence.measures(sonuc)
```

```
Influence measures of  
lm(formula = lny ~ x1 + x2 + x3 + x4) :  
  
      dfb.1_    dfb.x1    dfb.x2    dfb.x3    dfb.x42    dfb.x43    dffit cov.r    cook.d  
1  1.07e-01 -9.00e-02  8.49e-02 -2.78e-02 -6.01e-02 -6.72e-02  1.29e-01 1.087 2.78e-03  
2  3.40e-02 -2.21e-02  1.86e-02 -6.40e-04 -4.95e-02 -5.08e-02  7.24e-02 1.057 8.79e-04  
3  1.71e-02 -1.87e-02  6.60e-04  3.65e-02 -5.96e-02 -5.72e-02  8.89e-02 1.057 1.32e-03  
4 -9.54e-02  7.01e-02 -1.36e-01  1.35e-01 -8.67e-02 -6.78e-02  1.98e-01 1.084 6.56e-03  
5  2.17e-02 -3.26e-02  4.61e-03 -5.73e-03  2.73e-02  2.82e-02 -5.77e-02 1.076 5.59e-04  
6  6.05e-02 -3.12e-02  4.12e-02 -4.62e-02 -4.18e-02 -4.75e-02  8.37e-02 1.074 1.17e-03  
7 -5.72e-03  6.07e-03 -4.01e-03  4.77e-03 -5.28e-03 -4.89e-03  1.11e-02 1.086 2.05e-05  
8  2.42e-02  3.46e-02 -6.66e-02 -6.42e-02 -7.76e-02 -7.47e-02  1.71e-01 1.058 4.91e-03  
9  2.24e-02 -1.04e-01 -4.46e-02  9.71e-02  6.57e-02  8.10e-02 -2.02e-01 1.067 6.84e-03  
10 9.56e-02 -8.03e-02  1.69e-02 -1.73e-03 -8.90e-02 -8.74e-02  1.43e-01 1.048 3.42e-03  
11 2.50e-02  8.00e-02  1.36e-02  2.37e-02  5.62e-02  5.67e-02  0.60e-01 1.055 1.54e-03
```

13,20,22,52,60,63,68,80,90,107,119 ve 125.gözlem değerleri aykırı-uç değerlerdir. Veriden çıkarılarak işlemler tekrar edilecektir.12 gözlem silinmiş olup, 138 gözlemle analize devam edeceğiz.

Aykırı Değer İncelemesinden Sonraki Analiz Kısmı

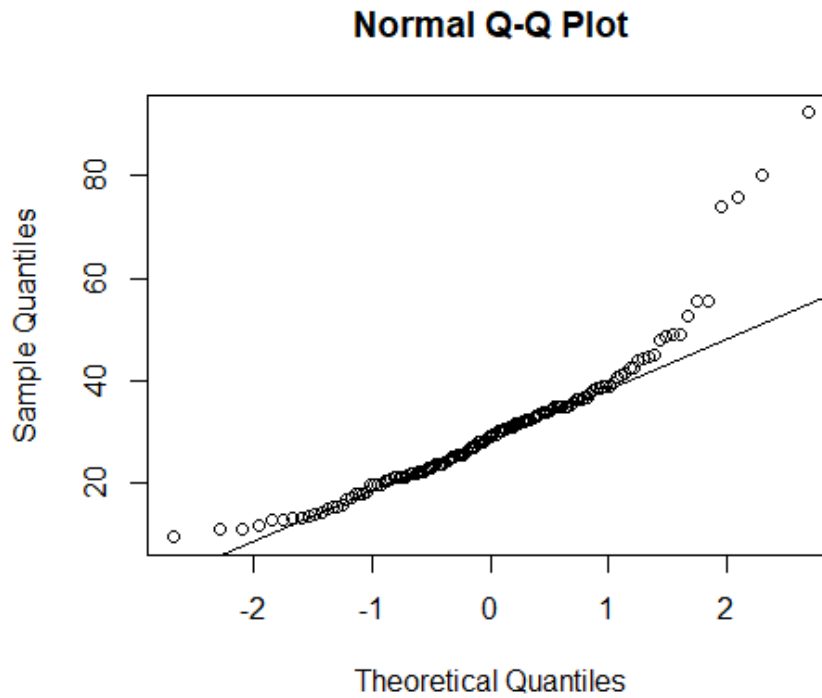
```
library(readxl)  
regresyon_yeni_verii <- read_excel("HOCAYA ATILACAKLAR/regresyon_yeni_verii.xlsx")  
view(regresyon_yeni_verii)  
names(regresyon_yeni_verii)<-c("y1","x1_","x2_","x3_","x4_")  
names(regresyon_yeni_verii)  
x4<-as.factor(regresyon_yeni_verii$x4_)  
class(regresyon_yeni_verii$x4_)  
attach(regresyon_yeni_verii)
```

```
install.packages("psych")  
library(psych)  
ozet_ıkı<-describe(regresyon_yeni_verii)  
view(ozet_ıkı)  
data.frame(ozet_ıkı)
```

	vars	n	mean	sd	median	trimmed	mad	min	max	range
y1	1	138	30.2892835	13.1202953	29.2315719	28.9511309	10.504770	9.5100645	92.197618	82.687553
x1_	2	138	6.0189576	1.0259053	6.0117557	6.0178362	1.023736	3.8969804	8.465775	4.568794
x2_	3	138	0.1211833	1.0054870	0.2113841	0.1249852	1.000601	-2.3727331	2.378632	4.751365
x3_	4	138	3.1395682	0.9743488	3.2257016	3.1464358	1.189196	0.4070959	5.404671	4.997575
x4_	5	138	2.0000000	0.8283305	2.0000000	2.0000000	1.482600	1.0000000	3.000000	2.000000
			skew	kurtosis	se					
y1			1.67519684	4.8494150	1.11687371					
x1_			0.04375241	-0.4624920	0.08733086					
x2_			-0.12585829	-0.6470312	0.08559274					
x3_			-0.11910672	-0.5765811	0.08294208					
x4_			0.00000000	-1.5531144	0.07051217					

Bağımlı değişken adını y1 olarak aldım bağımsız değişkenleri x1_, x2_, x3_, x4_ olarak aldım. Normalliğini tekrar inceleyelim.

```
install.packages("graphics")  
library(graphics)  
qqnorm(y1)  
qqline(y1)
```



Grafikte verilerin 45° doğru etrafında nasıl dağıldığını görüyoruz. Daha güvenilir olması için normallik testi yapılmalıdır.

```
> library(nortest)
> lillie.test(y1)

      Lilliefors (Kolmogorov-Smirnov) normality test

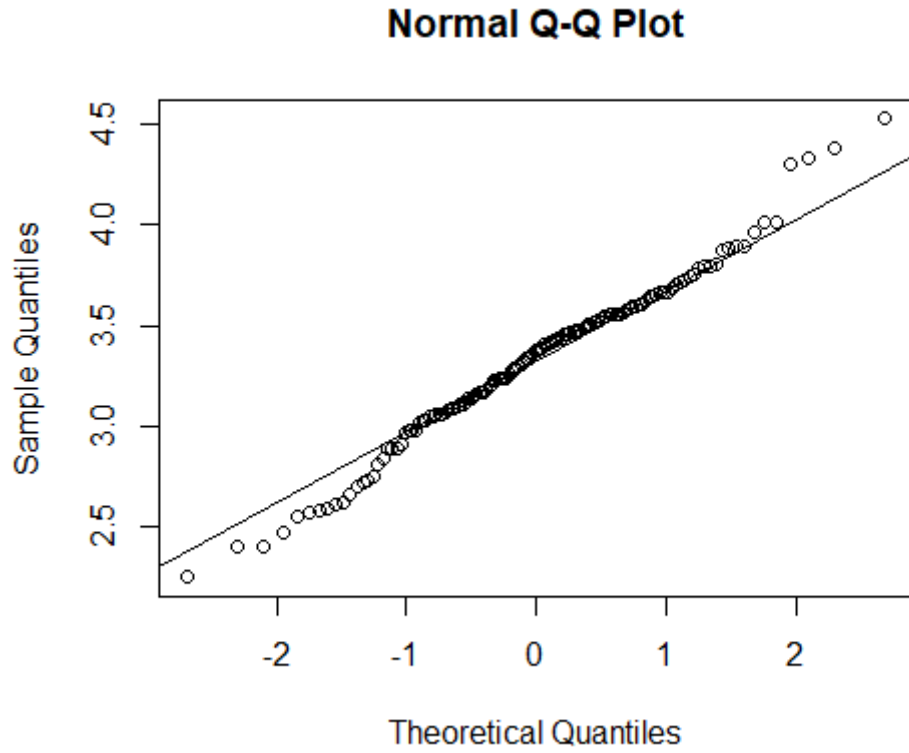
data:  y1
D = 0.10689, p-value = 0.0005545
```

Bağımlı değişkenimizin p değeri $=0.0005545 < 0.05$ olduğundan ,normal dağılıma uymamaktadır. Dönüşüm uygulanmalıdır. Dönüşüm uygulandıktan sonraki normal dağılım grafiği aşağıdaki gibidir.

```
> lny1<-log(y1)
> qqnorm(lny1)
> qqline(lny1)
> lillie.test(lny1)

      Lilliefors (Kolmogorov-Smirnov) normality test

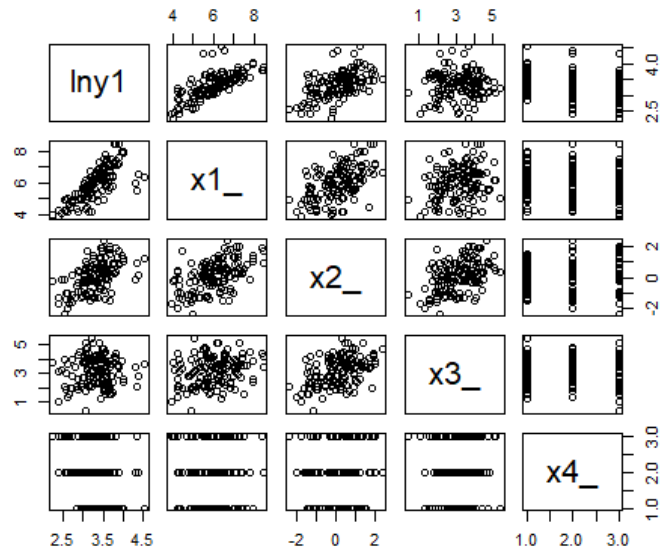
data:  lny1
D = 0.054933, p-value = 0.3903
```



Grafikte de görüldüğü gibi verilerin doğrunun etrafında dağıldığını söyleyebiliriz. Kolmogorov simirnov test sonucunda p değeri $=0.3903 > 0.05$ olup, y_1 bağımlı değişkenimizin normal dağılıma uyduğundan emin olabiliriz.

Grafikte bağımsız x_1 değişkeni arttıkça $\ln y_1$ bağımlı değişkeni arttığı görülmektedir. x_4 değişkeni nitel bir değişken olduğundan doğrusallık incelemesinde kullanılmamaktadır.

Bağımlı değişkenimizin normalliğini test ettikten sonra regresyon modelimizi kurup anlamlılığını test edebiliriz.



Kuracağımız yeni regresyon modelimizin model anlamlılığı incelemesinden önce, tekrar aykırı değer incelemesi yapacağız. Ödev gereği belirlenen aykırı değerlerden sonra bir değişiklik yapılmadan analize devam edilecektir.

```
sonuc_1k1<-lm(lm1~x1_+x2_+x3_+x4_)
summary(sonuc_1k1)

> regresyon_yeni_verii<-cbind(lm1,x1_,x2_,x3_,x4_)
> view(regresyon_yeni_verii)
> pairs(regresyon_yeni_verii)
> sonuc_1k1<-lm(lm1~x1_+x2_+x3_+x4_)
> summary(sonuc_1k1)

Call:
lm(formula = lm1 ~ x1_ + x2_ + x3_ + x4_)

Residuals:
    Min       1Q   Median       3Q      Max
-0.34220 -0.08623  0.00087  0.03727  1.11172

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.40794     0.14225   16.927 < 2e-16 ***
x1_          0.23447     0.02051   11.432 < 2e-16 ***
x2_          0.15023     0.02294    6.550 1.18e-09 ***
x3_         -0.10225     0.02069   -4.941 2.31e-06 ***
x4_2        -0.23993     0.04407   -5.444 2.45e-07 ***
x4_3        -0.32861     0.04332   -7.586 5.26e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2084 on 132 degrees of freedom
Multiple R-squared:  0.749,    Adjusted R-squared:  0.7395
F-statistic: 78.77 on 5 and 132 DF,  p-value: < 2.2e-16
```

İkinci Aykırı Değer İncelemesi

h_{ij} (Gözlem Uzaklığı:Uç Değerler)

```
predict(sonuc_1k1)
inff<-ls.diag(sonuc_1k1)
inff

> inff<-ls.diag(sonuc_1k1)
> inff
$std.dev
[1] 0.2083718

$hat
[1] 0.05625167 0.02386158 0.02724497 0.06504518 0.03810765 0.03881305 0.04304114 0.04570289
[9] 0.05784437 0.03270800 0.02810756 0.08616456 0.02224259 0.02487009 0.06479833 0.06282733
[17] 0.05955752 0.03346632 0.07680132 0.05457311 0.05244396 0.04221097 0.03446668 0.03629099
[25] 0.03465543 0.04954807 0.02894662 0.05455446 0.03498633 0.05417405 0.04720411 0.03853090
[33] 0.02921619 0.02651719 0.03719695 0.02779208 0.02767208 0.02509846 0.02329465 0.04965495
[41] 0.04988662 0.03550244 0.03903804 0.03466471 0.02445034 0.03537459 0.03210771 0.08542155
[49] 0.03566364 0.03769251 0.02486117 0.06599724 0.03623690 0.03913483 0.10119333 0.04465094
[57] 0.02388835 0.03566007 0.04058372 0.07249474 0.07245605 0.03620058 0.02804899 0.04390694
[65] 0.04705562 0.05186623 0.06696356 0.03891659 0.03308725 0.04102482 0.05161867 0.02863714
[73] 0.05280857 0.04639339 0.05723973 0.02952759 0.03543907 0.02554631 0.04658818 0.04416916
[81] 0.07736350 0.02831287 0.03137566 0.04068436 0.04294986 0.06671590 0.02878995 0.04603299
[89] 0.02918033 0.04733927 0.03823716 0.04942055 0.04259238 0.03276866 0.06469091 0.02891103
[97] 0.07960541 0.07442333 0.03482194 0.03945955 0.05804110 0.02651672 0.04319046 0.07400645
[105] 0.04661235 0.03136731 0.05334780 0.07473774 0.02695889 0.02609869 0.06026904 0.03761812
[113] 0.05096044 0.03177204 0.03515544 0.04384270 0.04567727 0.03026329 0.02259056 0.03106825
[121] 0.02312210 0.06527530 0.05298099 0.03890901 0.04533580 0.04224726 0.06819007 0.03915486
[129] 0.02970348 0.02407684 0.05688640 0.05278666 0.02365990 0.03023603 0.04024839 0.04031231
[137] 0.04701309 0.03760899
```


$h_{ii} > (2(k+1))/n$ olduğu durumda uç değer olur. ($k=4, n=138$) $h_{ii} = (2 \times 5)/138 = 0,0725$ değeridir. $h_{ii} = 0,0725$ değerinden yüksek değerler olarak, 19, 45, 60, 97, 98, 104 ve 108. Gözlemler $> 0,0725$ olduğundan uç değerlerdir.

Std.res(Aykırı Değerler)

```
$std.res
[1] 0.524963937 0.463016875 0.542917603 0.770014027 -0.286431063 0.396617183 0.070892853
[8] 0.760473086 -0.847456875 0.763704113 0.548229113 0.782173039 0.478668153 0.428734094
[15] 0.932571832 0.028921132 0.630760854 3.916141615 0.742719951 0.606954280 0.506229439
[22] 0.892005453 0.106126287 0.883512662 -0.328205510 0.852603392 0.133314942 -0.853116400
[29] -0.304393871 0.761268452 -0.626900453 0.365898026 0.559477676 0.155245687 -1.094188476
[36] -1.204023432 -0.991290717 -1.125099311 -1.080935994 -1.013041201 -1.459134976 -1.275003920
[43] -1.535193727 -1.152233836 -1.001161295 -1.141271494 -1.210338252 0.347189798 -0.064821816
[50] 0.166241804 0.045588833 -0.563017810 0.189760573 -0.199259513 -0.123599272 -0.232611364
[57] 5.400184001 0.003421061 -0.025102563 -0.891963000 -0.414304723 -0.021679133 0.022946575
[64] -0.104132816 -0.040404915 -0.413842364 -0.280488600 0.057438871 0.033867576 0.255285572
[71] -0.869044139 4.595120638 0.084724581 -1.192945447 -0.099039939 -0.077082127 -0.415386792
[78] 0.043012096 -0.440866461 -0.214297250 0.149207100 -0.298629300 -0.367929918 -1.053409354
[85] -0.428441458 -0.575270478 -0.421206572 -0.733402239 -0.421100443 -0.109000403 -0.418594427
[92] -0.692976270 0.116192013 0.152640117 0.201943121 0.138040001 0.212089232 -1.069785091
[99] 0.059209885 0.091970210 0.077800505 0.231566793 -0.063874147 -0.617994172 0.090296263
[106] -0.083111362 0.359660421 -0.414189721 0.049887111 0.047175074 -1.694100994 0.005116704
[113] 5.084824808 -0.090513739 0.133916557 -0.057599141 -0.596719649 0.015878105 0.114453026
[120] 0.076528194 0.141065655 -0.532353761 0.067756383 -0.075302523 -0.307518088 0.309656122
[127] -0.241157500 0.086823614 0.186939015 0.130817999 -0.415134207 -1.071911623 0.149425677
[134] 0.112954783 0.043544178 -0.438703747 -0.031891717 -0.046480622
```

Standartlaştırılmış hataların (-2,+2) aralığında olması istenir. Bu aralığa girmeyen gözlemler: 18, 57, 72 ve 113. gözlemlerdir. Aykırı değerler olarak belirlenir.

Stud.res (Student Tipi Artıklar= r_i)

```
$stud.res
[1] 0.523518438 0.461634718 0.541462075 0.768820411 -0.285432751 0.395347629 0.070625153
[8] 0.759252076 -0.846546787 0.762492201 0.546771375 0.781016662 0.477265963 0.427404703
[15] 0.932108349 0.028811465 0.629316188 4.149788015 0.741452169 0.605496358 0.504798510
[22] 0.891310616 0.105728039 0.882773715 -0.327093434 0.851716159 0.132817943 -0.852231474
[29] -0.303345155 0.760049664 -0.625453092 0.364694407 0.558016426 0.154670639 -1.095013174
[36] -1.206095241 -0.991225106 -1.126242677 -1.081631528 -1.013142723 -1.465463981 -1.278059476
[43] -1.543206333 -1.153677465 -1.001170176 -1.142591472 -1.212491730 0.346030216 -0.064576840
[50] 0.165628242 0.045416177 -0.561555787 0.189066206 -0.198533168 -0.123137329 -0.231776094
[57] 6.094910993 0.003408078 -0.025007356 -0.891267939 -0.413001017 -0.021596897 0.022859537
[64] -0.103741885 -0.040251824 -0.412539514 -0.279507431 0.057221601 0.033739192 0.254379547
[71] -0.868233407 4.994550709 0.084405339 -1.194876683 -0.098667741 -0.076791323 -0.414081091
[78] 0.042849162 -0.439517040 -0.213521120 0.148653383 -0.297596524 -0.366721688 -1.053850581
[85] -0.427112567 -0.573807025 -0.419890332 -0.732112061 -0.419784392 -0.108591625 -0.417282877
[92] -0.691605546 0.115756975 0.152074257 0.201207815 0.137526054 0.211320345 -1.070375350
[99] 0.058985962 0.091624111 0.077507023 0.230734850 -0.063632722 -0.616541403 0.089956360
[106] -0.082798114 0.358471169 -0.412886228 0.049698254 0.046996437 -1.706323346 0.005097286
[113] 5.648885585 -0.090173031 0.133417396 -0.057381269 -0.595258457 0.015817861 0.114024326
[120] 0.076239455 0.140540894 -0.530903661 0.067500416 -0.075018355 -0.306460829 0.308593056
[127] -0.240295229 0.086496581 0.186254223 0.130329983 -0.413828971 -1.072521719 0.148871186
[134] 0.112531549 0.043379236 -0.437357789 -0.031770808 -0.046304604
```

Student tipi artıkların (-3,+3) aralığında olması istenen durumdur. Bu aralığa girmeyen gözlemler: 18, 57, 72 ve 113. gözlemlerdir. Aykırı değerler olarak belirlenir.

Cook Uzaklığı(D_i)

```
$cooks
[1] 2.737706e-03 8.734340e-04 1.375940e-03 6.874965e-03 5.417197e-04 1.058673e-03 3.767421e-05
[8] 4.616114e-03 7.348901e-03 3.286968e-03 1.448698e-03 9.614244e-03 8.687046e-04 7.813393e-04
[15] 1.004320e-02 9.345625e-06 4.199355e-03 8.850270e-02 7.648432e-03 3.544145e-03 2.363927e-03
[22] 5.844391e-03 6.700803e-05 4.899223e-03 6.445097e-04 6.315961e-03 8.830009e-05 6.999372e-03
[29] 5.598678e-04 5.532284e-03 3.245084e-03 8.942166e-04 1.570056e-03 1.094176e-04 7.709086e-03
[36] 6.906858e-03 4.661008e-03 5.431463e-03 4.644526e-03 8.936845e-03 1.863152e-02 9.973069e-03
[43] 1.595721e-02 7.945831e-03 4.186898e-03 7.960850e-03 8.099246e-03 1.876416e-03 2.589931e-05
[50] 1.804141e-04 8.831221e-06 3.733109e-03 2.256532e-04 2.695181e-04 2.866594e-04 4.214820e-04
[57] 1.189468e-01 7.213107e-08 4.442522e-06 1.036412e-02 2.234749e-03 2.942127e-06 2.532546e-06
[64] 8.299597e-05 1.343572e-05 1.561470e-03 9.410638e-04 2.226559e-05 6.541697e-06 4.646657e-04
[71] 6.851036e-03 1.037506e-01 6.670129e-05 1.153923e-02 9.925805e-05 3.013012e-05 1.056591e-03
[78] 8.083452e-06 1.582917e-03 3.536877e-04 3.111238e-04 4.330828e-04 7.308302e-04 7.843486e-03
[85] 1.372963e-03 3.942833e-03 8.765299e-04 4.325826e-03 8.883247e-04 9.839845e-05 1.161056e-03
[92] 4.161066e-03 1.001005e-04 1.315572e-04 4.701051e-04 9.455035e-05 6.484172e-04 1.533694e-02
[99] 2.108059e-05 5.791347e-05 6.216090e-05 2.434404e-04 3.069455e-05 5.087204e-03 6.643849e-05
[106] 3.728102e-05 1.214955e-03 2.309524e-03 1.149202e-05 9.939800e-06 3.067737e-02 1.705607e-07
[113] 2.313927e-01 4.480693e-05 1.089062e-04 2.535413e-05 2.840496e-03 1.311319e-06 5.046077e-05
[120] 3.129791e-05 7.850156e-05 3.298486e-03 4.280658e-05 3.826075e-05 7.484808e-04 7.049417e-04
[127] 7.093231e-04 5.119844e-05 1.783001e-04 7.036677e-05 1.732488e-03 1.067193e-02 9.018017e-05
[134] 6.630050e-05 1.325252e-05 1.347408e-03 8.362504e-06 1.407125e-05
```

$n > 50$ olduğu durumlarda cook uzaklığı(D_i)değeri $4/n$ 'den büyük ise etkili gözlem olduğu söylenir.($n=141$) $\rightarrow 4/138 = 0,0290$ değerinden büyük olan 18 ve 111.gözlemlerdir.Bu gözlemler aykırı ve aynı zamanda etkin gözlemlerdir.

Belirlenen uç-aykırı değerler: 18,19,45,57,60,72,97,98 ,104,108,111,113

Şimdi değişiklik yapmadan regresyon modelinin anlamlılığını, katsayı yorumlarını ve güven aralıklarını inceleyeceğiz.

```
call:
lm(formula = lny1 ~ x1_ + x2_ + x3_ + x4_)

Residuals:
    Min       1Q   Median       3Q      Max
-0.34220 -0.08623  0.00087  0.03727  1.11172

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.40794    0.14225   16.927  < 2e-16 ***
x1_          0.23447    0.02051   11.432  < 2e-16 ***
x2_          0.15023    0.02294    6.550 1.18e-09 ***
x3_         -0.10225    0.02069   -4.941 2.31e-06 ***
x4_2         -0.23993    0.04407   -5.444 2.45e-07 ***
x4_3         -0.32861    0.04332   -7.586 5.26e-12 ***
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2084 on 132 degrees of freedom
Multiple R-squared:  0.749,    Adjusted R-squared:  0.7395
F-statistic: 78.77 on 5 and 132 DF,  p-value: < 2.2e-16
```

Regresyon denklemi:

$$y_1 = 2.40794 + 0.23447x_{1_} + 0.15023x_{2_} - 0.10225x_{3_} - 0.23993x_{4_2} - 0.32861x_{4_3} \pm 0.2084$$
$$(0.14225) \quad (0.02051) \quad (0.02294) \quad (0.02069) \quad (0.04407) \quad (0.04332)$$

Model Anlamlılığı:

$H_0 : \beta = 0$ (Model anlamlı değildir)

$H_1 : \beta \neq 0$ (Model anlamlıdır)

$P = 0,00 < \alpha = 0,05$ H_0 ret.

$P = 0.00$ ($2.2e-16$) < 0.05 olduğundan modelin anlamlı olduğu %95 güven düzeyinde söylenir.

Katsayı Yorumları:

$b_0 = 2.40794$; denklem gereği katsayıdır.

$b_1 = 0.23447$; Diğer değişkenler sabit tutulduğunda, x_1 bağımsız değişkenindeki bir birimlik artış y_1 bağımlı değişkeninde ortalama 0.23447 artış sağlar. $P = 0,00 < 0,05$ olduğundan katsayı istatistiksel olarak anlamlıdır.

$b_2 = 0.15023$; Diğer değişkenler sabit tutulduğunda, x_2 bağımsız değişkenindeki bir birimlik artış y_1 bağımlı değişkeninde ortalama 0.15023 artış sağlar. $P = 0,00 < 0,05$ olduğundan katsayı istatistiksel olarak anlamlıdır.

$b_3 = -0.10225$; Diğer değişkenler sabit tutulduğunda, x_3 bağımsız değişkenindeki bir birimlik artış y_1 bağımlı değişkeninde ortalama 0.10225 azalış sağlar. $P = 0,00 < 0,05$ olduğundan katsayı istatistiksel olarak anlamlıdır.

$b_{4_2} = -0.23993$; Diğer değişkenler sabit tutulduğunda, x_3 bağımsız değişkenindeki bir birimlik artış y_1 bağımlı değişkeninde ortalama 0.23993 azalış sağlar. $P = 0,00 < 0,05$ olduğundan katsayı istatistiksel olarak anlamlıdır.

$b_{4_3} = -0.32861$; Diğer değişkenler sabit tutulduğunda, x_3 bağımsız değişkenindeki bir birimlik artış y_1 bağımlı değişkeninde ortalama 0.32861 azalış sağlar. $P = 0,00 < 0,05$ olduğundan katsayı istatistiksel olarak anlamlıdır.

Belirtme Katsayısı:

Belirtme katsayısı bağımlı değişkendeki değişimin yüzde kaçının bağımsız değişken tarafından açıklandığını gösterir. Multiple R squared: $0.749 = R^2$, Bağımlı değişken y_1 'in %75'lik kısmı bağımsız değişkenler (x_1, x_2, x_3, x_4) ile açıklanmaktadır. Açıklanamayan kısım %25'dir. Gözlem sayısının artırılması ya da açıklanan değişkeni anlamlı olarak açıklayabilecek başka değişkenlerin modele eklenmesi belirtme katsayısını artırır.

Bağımsız değişkenlerin (x_1, x_2, x_3, x_4) bağımlı değişken y_1 ile birlikte ilişkisi de %74'tür.

%99 Güven Düzeyleri:

```
> confint(sonuc_1k1, level = .99)
              0.5 %      99.5 %
(Intercept)  2.03614693  2.77973308
x1_          0.18086879  0.28807966
x2_          0.09028632  0.21018044
x3_         -0.15632927 -0.04816392
x4_2         -0.35511550 -0.12474101
x4_3         -0.44182326 -0.21539784
```

Bağımsız değişkenlerin katsayı kestirimlerinin 0 içerip içermediğine bakarak yorumumuzu yapacağız.

$\beta_0 \rightarrow$ için güven aralığı : $P(2.0361 \leq \beta \leq 2.7797) = 0.99 \Rightarrow 0$ 'ı içermiyor.

$\beta_1 \rightarrow$ için güven aralığı : $P(0.1809 \leq \beta \leq 0.2881) = 0.99 \Rightarrow 0$ 'ı içermiyor. x_1 bağımsız değişkeninin bağımlı y_1 değişkenini açıklamada etkili olduğu %99 güvenirlikle söylenebilir.

$\beta_2 \rightarrow$ için güven aralığı : $P(0.0903 \leq \beta \leq 0.2102) = 0.99 \Rightarrow 0$ 'ı içermiyor. x_2 bağımsız değişkeninin bağımlı y_1 değişkenini açıklamada etkili olduğu %99 güvenirlikle söylenebilir.

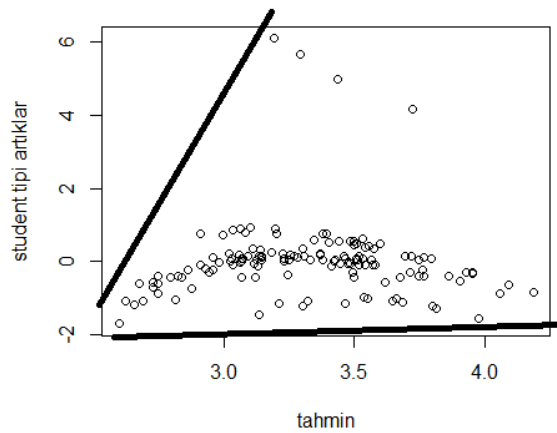
$\beta_3 \rightarrow$ için güven aralığı : $P(-0.1563 \leq \beta \leq -0.1247) = 0.99 \Rightarrow 0$ 'ı içermiyor. x_3 bağımsız değişkeninin bağımlı y_1 değişkenini açıklamada etkili olduğu %99 güvenirlikle söylenebilir.

$\beta_{4,2} \rightarrow$ için güven aralığı : $P(-0.3551 \leq \beta \leq 0.1093) = 0.99 \Rightarrow 0$ 'ı içermiyor. $x_{4,2}$ bağımsız nitel değişkeninin bağımlı y_1 değişkenini açıklamada etkili olduğu %99 güvenirlikle söylenebilir.

$\beta_{4,3} \rightarrow$ için güven aralığı : $P(-0.4418 \leq \beta \leq 0.2154) = 0.99 \Rightarrow 0$ 'ı içermiyor. x_4 bağımsız nitel değişkeninin bağımlı y_1 değişkenini açıklamada etkili olduğu %99 güvenirlikle söylenebilir.

Değişen Varyanslık:

```
par(mfrow=c(1,1))
plot(predict(sonuc_1k1), inff$stud.res, ylab="student tipi artıklar", xlab="tahmin")
```



Student tipi artıklarla kestirim değerleri arasında çizilmiş olan grafik incelendiğinde rasgele olmadığı, gözlemlerin nokta olarak ifade edildiği durumda sağa doğru megafon olduğu görülmektedir. Şimdi Bruege-Pagan testini uygulayacağız.

Bruege-Pagan

```
library(lmtest)
bptest(sonuc_1k1)
```

studentized Breusch-Pagan test

```
data: sonuc_1k1
BP = 1.8021, df = 5, p-value = 0.8758
```

H_0 =Varyanslar homojendir.

Test sonucunda $p=0.8758>0.05$ bulunmuştur değişen varyanslılık olmadığı %95 güven düzeyinde söylenir. Ayrıca White testi ile de test edilebilir.

White testi

```
library(lmtest)
res=residuals(sonuc_1k1)
sqres=res^2
sqx1=x1_*x1_
sqx2=x2_*x2_
sqx3=x3_*x3_
x1x2=x1_*x2_
x1x3=x1_*x3_
x2x3=x2_*x3_
WH=lm(sqres ~ x1_+x2_+x3_+sqx1+sqx2+sqx3+x1x2+x1x3+x2x3)
whs= summary(WH)
whts=whs$r.squared*length(WH$residuals)
MHPv=1-pchisq(whts,df=WH$rank-1)
MHPv
whts

> library(lmtest)
> res=residuals(sonuc_1k1)
> sqres=res^2
> sqx1=x1_*x1_
> sqx2=x2_*x2_
> sqx3=x3_*x3_
> x1x2=x1_*x2_
> x1x3=x1_*x3_
> x2x3=x2_*x3_
> WH=lm(sqres ~ x1_+x2_+x3_+sqx1+sqx2+sqx3+x1x2+x1x3+x2x3)
> whs= summary(WH)
> whts=whs$r.squared*length(WH$residuals)
> MHPv=1-pchisq(whts,df=WH$rank-1)
> MHPv
[1] 0.8891332
> whts
[1] 4.319787
```

Öz İlişki Sorunu

```
library(lmtest)
dwtest(sonuc_1k1)

> dwtest(sonuc_1k1)

Durbin-Watson test

data: sonuc_1k1
DW = 1.7071, p-value = 0.02866
alternative hypothesis: true autocorrelation is greater than 0
```

H_0 =Öz ilişki yoktur.

H_s =Pozitif öz ilişki vardır.

Durbin test istatistiği (DW=1.6763) sonucunda, alternatif hipotez kurulurken 0-2 aralığında yer aldığından pozitif öz ilişki vardır şeklinde kurulur. ($P=0.02866<0.05$) olduğundan h_0 reddedilir. Pozitif öz ilişki olduğu görülmektedir.

Çoklu Bağlantı Sorunu İncelemesi

```
$correlation
(Intercept)      x1_      x2_      x3_      x4_2      x4_3
(Intercept)  1.0000000 -0.872234061  0.59314627 -0.434133198 -0.17179521 -0.230109184
x1_          -0.8722341  1.000000000 -0.48183503 -0.006586504  0.07845408  0.087637119
x2_          0.5931463 -0.481835027  1.000000000 -0.402948173  0.01293618 -0.113718568
x3_          -0.4341332 -0.006586504 -0.40294817  1.000000000 -0.10072391  0.009416027
x4_2         -0.1717952  0.078454081  0.01293618 -0.100723909  1.000000000  0.488650188
x4_3         -0.2301092  0.087637119 -0.11371857  0.009416027  0.48865019  1.000000000
```

Korelasyon tablosu incelendiğinden değişkenler arasında güçlü ilişki olmadığı görülmektedir.

```
install.packages("car")
library(car)
install.packages("DAAG")
library(DAAG)
vif(sonuc_1k1)

> vif(sonuc_1k1)
      x1_      x2_      x3_      x4_2      x4_3
1.3970 1.6783 1.2827 1.3408 1.3394
```

Çoklu bağlantıdan şüphelenmemiz için VIF değerlerinin çok büyük olması gerekir. VIF değerleri birbirlerine çok yakın, bu yüzden şüphelenmemize gerek yok.

Not: "perturb" paketi birçok sürüm denememe rağmen indirilmedi. Bu yüzden koşul sayısının 30'dan büyük olması durumunu inceleyemedim.

Özdeğerler ve özvektörler

```
install.packages("fastDummies")
library(fastDummies)
dummy<-dummy_cols(x4_)
x41<-dummy$.data_1
x42<-dummy$.data_2
x43<-dummy$.data_3
ort1<-mean(x1_)
kt1<-sum((x1_-ort1)^2)
skx1<-(x1_-ort1)/(kt1^0.5)
ort2<-mean(x2_)
kt2<-sum((x2_-ort2)^2)
skx2<-(x2_-ort2)/(kt2^0.5)
ort3<-mean(x3_)
kt3<-sum((x3_-ort3)^2)
skx3<-(x3_-ort3)/(kt3^0.5)
ort42<-mean(x42)
kt42<-sum((x42-ort42)^2)
skx42<-(x42-ort42)/(kt42^0.5)
ort43<-mean(x43)
kt43<-sum((x43-ort43)^2)
skx43<-(x43-ort43)/(kt43^0.5)
x<-cbind(skx1,skx2,skx3,skx42,skx43)
sm<- eigen (t(x)%*%x)
signif(sm$values,3)
signif(sm$vectors,3)

|

> signif(sm$values,3)
[1] 1.850 1.490 0.765 0.500 0.397
> signif(sm$vectors,3)
      [,1]      [,2]      [,3]      [,4]      [,5]
[1,] -0.552 -0.0868  0.6460  0.0105  0.520
[2,] -0.640 -0.0733  0.0344  0.1990 -0.738
[3,] -0.489 -0.2340 -0.7160 -0.2810  0.338
[4,]  0.143 -0.6900 -0.0839  0.6930  0.127
[5,] -0.164  0.6760 -0.2470  0.6330  0.234
> |
```

signif(sm\$value,3) kodunun altındaki değerlere bakıldığında, özdeğerlerde sıfıra çok yakın sayı bulunmamaktadır. Çoklu bağlantı sorunu olmadığı söylenebilir. Bu değerlerden sıfıra çok yakın olanları, bize kaç tane güçlü çoklu bağlantı olduğunu gösterir.

```
> V<-sm$Vectors
> t(V)%*%V
      [,1]      [,2]      [,3]      [,4]      [,5]
[1,] 1.000000e+00 0.000000e+00 2.775558e-17 5.551115e-17 -7.632783e-17
[2,] 0.000000e+00 1.000000e+00 2.775558e-17 0.000000e+00 0.000000e+00
[3,] 2.775558e-17 2.775558e-17 1.000000e+00 5.551115e-17 -4.857226e-16
[4,] 5.551115e-17 0.000000e+00 5.551115e-17 1.000000e+00 8.326673e-17
[5,] -7.632783e-17 0.000000e+00 -4.857226e-16 8.326673e-17 1.000000e+00
> |
```

Özdeğerler ve özvektörlerin ilişki matrisi aşağıdaki gibidir.

```
> V %*% diag(sm$values) %*% t(V)
      [,1]      [,2]      [,3]      [,4]      [,5]
[1,] 1.000000000 0.52671433 0.24256118 -0.06798352 0.008547327
[2,] 0.526714327 1.00000000 0.45691150 -0.06401471 0.107239253
[3,] 0.242561176 0.45691150 1.00000000 0.07773876 -0.010769308
[4,] -0.067983518 -0.06401471 0.07773876 1.00000000 -0.491689172
[5,] 0.008547327 0.10723925 -0.01076931 -0.49168917 1.000000000
> |
```

Uyum kestirimi

$$y_1^{\sim} = 2.40794 + 0.23447x_{1_} + 0.15023x_{2_} - 0.10225x_{3_} - 0.23993x_{4_2} - 0.32861x_{4_3} \pm 0.2084$$

$$(0.14225) \quad (0.02051) \quad (0.02294) \quad (0.02069) \quad (0.04407) \quad (0.04332)$$

Uyum kestirimi için 52.gözlem seçilmiştir.

$$(x_{1_}=6,268267, x_{2_}=0,509594, x_{3_}=3,277024, x_{4_2}=1, x_{4_3}=0)$$

Katsayılar denklemde yerine konulduğunda uyum kestirimi 3,17081 olarak bulunur.

Ön Kestirim

$$y_1^{\sim} = 2.40794 + 0.23447x_{1_} + 0.15023x_{2_} - 0.10225x_{3_} - 0.23993x_{4_2} - 0.32861x_{4_3} \pm 0.2084$$

$$(0.14225) \quad (0.02051) \quad (0.02294) \quad (0.02069) \quad (0.04407) \quad (0.04332)$$

Ön kestirim için veride bulunmayan gözlem değerleri kullanılmaktadır.

$$(x_{1_}=7,234567, x_{2_}=1,548594, x_{3_}=1,012024, x_{4_2}=1, x_{4_3}=0)$$

Katsayılar denklemde yerine konulduğunda ön kestirim 3,7851 olarak bulunur.

Değişken Seçimi Yöntemi

- İleriye Doğru Adım Yöntemi

```
library(stats)
lm.null <- lm(lny1 ~ 1)
forward <- step(lm.null, lny1~x1_+x2_+x3_+x4_, direction = "forward")
summary(forward)
```

```
> library(stats)
> lm.null <- lm(lny1 ~ 1)
> forward <- step(lm.null, lny1~x1_+x2_+x3_+x4_, direction = "forward")
Start: AIC=-246.28
lny1 ~ 1
```

	Df	Sum of Sq	RSS	AIC
+ x1_	1	12.5149	10.316	-353.90
+ x2_	1	6.7922	16.039	-293.01
+ x4_	2	3.0772	19.754	-262.26
<none>			22.831	-246.28
+ x3_	1	0.0587	22.773	-244.63

```
Step: AIC=-353.9
lny1 ~ x1_
```

	Df	Sum of Sq	RSS	AIC
+ x4_	2	2.44783	7.8687	-387.28
+ x2_	1	0.76371	9.5528	-362.52
+ x3_	1	0.40288	9.9136	-357.40
<none>			10.3165	-353.90

```
Step: AIC=-387.28
lny1 ~ x1_ + x4_
```

	Df	Sum of Sq	RSS	AIC
+ x2_	1	1.07731	6.7914	-405.60
+ x3_	1	0.27465	7.5940	-390.19
<none>			7.8687	-387.28

```
Step: AIC=-405.6
lny1 ~ x1_ + x4_ + x2_
```

	Df	Sum of Sq	RSS	AIC
+ x3_	1	1.0601	5.7313	-427.02
<none>			6.7914	-405.60

```
Step: AIC=-427.02
```

```
Step: AIC=-427.02
lny1 ~ x1_ + x4_ + x2_ + x3_
```

```
> forward
```

```
Call:
lm(formula = lny1 ~ x1_ + x4_ + x2_ + x3_)
```

```
Coefficients:
```

(Intercept)	x1_	x4_2	x4_3	x2_	x3_
2.4079	0.2345	-0.2399	-0.3286	0.1502	-0.1022

Bağımlı değişken lny1 olduğu durumda, birinci adımda modele $x_{1_}$ değişkeni girmiştir. Sonrasında sırasıyla $x_{4_}$, $x_{2_}$, $x_{3_}$ değişkenleri eklenmiştir. Diğer değişkenler modele alındığında tüm değişkenler anlamlı çıktığından tüm değişkenler modelde yer almaktadır.

```
> summary(forward)

Call:
lm(formula = lny1 ~ x1_ + x4_ + x2_ + x3_)

Residuals:
    Min       1Q   Median       3Q      Max
-0.34220 -0.08623  0.00087  0.03727  1.11172

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.40794    0.14225   16.927 < 2e-16 ***
x1_          0.23447    0.02051   11.432 < 2e-16 ***
x4_2        -0.23993    0.04407   -5.444 2.45e-07 ***
x4_3        -0.32861    0.04332   -7.586 5.26e-12 ***
x2_          0.15023    0.02294    6.550 1.18e-09 ***
x3_         -0.10225    0.02069   -4.941 2.31e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2084 on 132 degrees of freedom
Multiple R-squared:  0.749,    Adjusted R-squared:  0.7395
F-statistic: 78.77 on 5 and 132 DF,  p-value: < 2.2e-16
```

Model anlamlılığı incelenirken $p=0.000$ olduğundan $p < \alpha = 0.05$ ile modelin anlamlı olduğu %95 güven ile söylenebilir Tüm değişkenler de $p=0.000 < \alpha = 0.05$ ile anlamlıdır.

En iyi model:

$$y_{1i} = 2.40794 + 0.23447x_{1_i} - 0.23993x_{4_2} - 0.32861x_{4_3} + 0.15023x_{2_i} - 0.10225x_{3_i} \pm 0.2084$$

(0.14225) (0.02051) (0.04407) (0.04332) (0.02294) (0.02069)

şeklinde kurulur.

- Geriye Doğru Adım Yöntemi

```
backward<-step(sonuc_1k1,direction="backward")
summary(backward)
```

```
Start:  AIC=-427.02
lmy1 ~ x1_ + x2_ + x3_ + x4_

      Df Sum of Sq  RSS   AIC
<none>      0      5.7313 -427.02
- x3_      1      1.0601  6.7914 -405.60
- x2_      1      1.8627  7.5940 -390.19
- x4_      2      2.6708  8.4021 -378.23
- x1_      1      5.6745 11.4057 -334.05
```

Modele ileriye doğru seçim yönteminin tam tersi olarak sırasıyla $x_{3_}$, $x_{2_}$, $x_{4_}$, $x_{1_}$ değişkenleri modele girmiştir. Anlamsız değişken olmadığından modelden değişken çıkarılmamıştır. Tüm değişkenler modelde yer almaktadır.

```
> summary(backward)
```

```
Call:
```

```
lm(formula = lny1 ~ x1_ + x2_ + x3_ + x4_)
```

```
Residuals:
```

```
      Min       1Q   Median       3Q      Max
-0.34220 -0.08623  0.00087  0.03727  1.11172
```

```
Coefficients:
```

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   2.40794     0.14225   16.927 < 2e-16 ***
x1_            0.23447     0.02051   11.432 < 2e-16 ***
x2_            0.15023     0.02294    6.550 1.18e-09 ***
x3_           -0.10225     0.02069   -4.941 2.31e-06 ***
x4_2           -0.23993     0.04407   -5.444 2.45e-07 ***
x4_3           -0.32861     0.04332   -7.586 5.26e-12 ***
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.2084 on 132 degrees of freedom
```

```
Multiple R-squared:  0.749,    Adjusted R-squared:  0.7395
```

```
F-statistic: 78.77 on 5 and 132 DF,  p-value: < 2.2e-16
```

Bu sonuçlarda da modelin ve katsayıların anlamlı olduğu %95 güvenle söylenir söylenir.

($p=0.000 < \alpha=0.05$)

- Adımsal Seçim Yöntemi

```
install.packages("MASS")
```

```
library(MASS)
```

```
step.model <- stepAIC(sonuc_1k1, direction = "both", trace = FALSE)
```

```
summary(step.model)
```

```
Call:
```

```
lm(formula = lny1 ~ x1_ + x2_ + x3_ + x4_)
```

```
Residuals:
```

```
      Min       1Q   Median       3Q      Max
-0.34220 -0.08623  0.00087  0.03727  1.11172
```

```
Coefficients:
```

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   2.40794     0.14225   16.927 < 2e-16 ***
x1_            0.23447     0.02051   11.432 < 2e-16 ***
x2_            0.15023     0.02294    6.550 1.18e-09 ***
x3_           -0.10225     0.02069   -4.941 2.31e-06 ***
x4_2           -0.23993     0.04407   -5.444 2.45e-07 ***
x4_3           -0.32861     0.04332   -7.586 5.26e-12 ***
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.2084 on 132 degrees of freedom
```

```
Multiple R-squared:  0.749,    Adjusted R-squared:  0.7395
```

```
F-statistic: 78.77 on 5 and 132 DF,  p-value: < 2.2e-16
```

En iyi model diğer modellerde de olduğu gibi tüm değişkenlerin olduğu modeldir.

Ridge Regresyon ve İz Grafiği

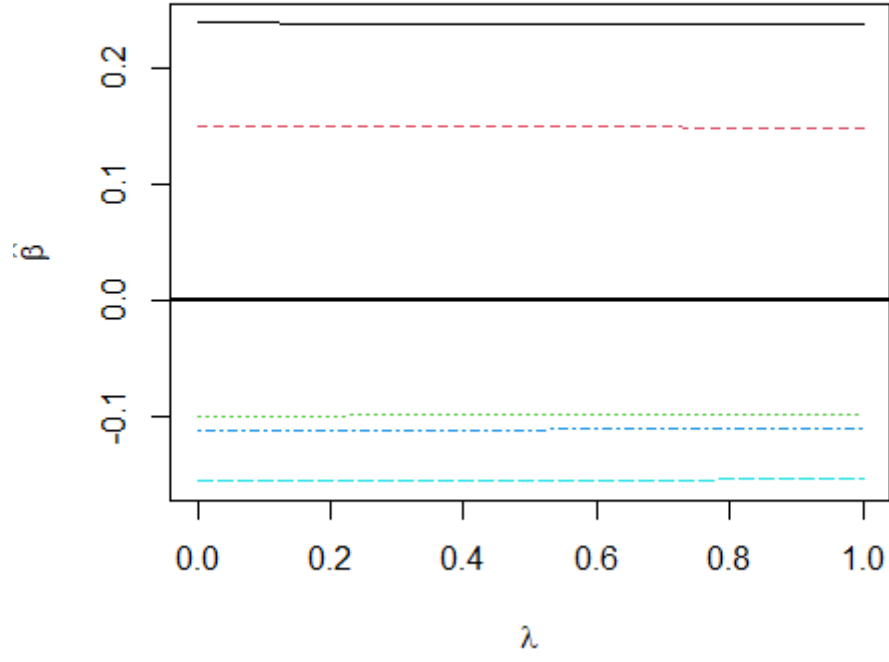
```
library(MASS)
ridge <- lm.ridge(lm1~x1_+x2_+x3_+x4_ ,lambda = seq(0,1,0.05))
matplot(ridge$lambda,t(ridge$coef),type="l",xlab=expression(lambda),
        ylab=expression(hat(beta)))
abline(h=0,lwd=2)
ridge$coef
select(ridge)
ridge$coef[,ridge$lam == 0.9]
```

```
> library(MASS)
> ridge <- lm.ridge(lm1~x1_+x2_+x3_+x4_ ,lambda = seq(0,1,0.05))
> matplot(ridge$lambda,t(ridge$coef),type="l",xlab=expression(lambda),
+       ylab=expression(hat(beta)))
> abline(h=0,lwd=2)
> ridge$coef
      0.00      0.05      0.10      0.15      0.20      0.25      0.30      0.35
x1_  0.23967521 0.2396049 0.23953469 0.23946447 0.23939427 0.23932411 0.23925397 0.23918385
x2_  0.15050941 0.1504519 0.15039447 0.15033711 0.15027983 0.15022262 0.15016548 0.15010843
x3_ -0.09926224 -0.0991878 -0.09911345 -0.09903919 -0.09896502 -0.09889093 -0.09881693 -0.09874302
x4_2 -0.11181307 -0.1117370 -0.11166102 -0.11158516 -0.11150941 -0.11143376 -0.11135822 -0.11128279
x4_3 -0.15572991 -0.1556285 -0.15552732 -0.15542623 -0.15532527 -0.15522445 -0.15512376 -0.15502321
      0.40      0.45      0.50      0.55      0.60      0.65      0.70      0.75
x1_  0.23911376 0.23904370 0.2389737 0.23890366 0.23883368 0.23876372 0.23869379 0.23862389
x2_  0.15005144 0.14999453 0.1499377 0.14988093 0.14982424 0.14976763 0.14971109 0.14965462
x3_ -0.09866919 -0.09859545 -0.0985218 -0.09844823 -0.09837475 -0.09830136 -0.09822805 -0.09815483
x4_2 -0.11120747 -0.11113225 -0.1110571 -0.11098214 -0.11090724 -0.11083244 -0.11075776 -0.11068318
x4_3 -0.15492279 -0.15482251 -0.1547224 -0.15462234 -0.15452246 -0.15442271 -0.15432309 -0.15422360
      0.80      0.85      0.90      0.95      1.00
x1_  0.23855402 0.23848417 0.23841435 0.2383446 0.2382748
x2_  0.14959822 0.14954190 0.14948565 0.1494295 0.1493734
x3_ -0.09808169 -0.09800864 -0.09793568 -0.0978628 -0.0977900
x4_2 -0.11060870 -0.11053433 -0.11046006 -0.1103859 -0.1103118
x4_3 -0.15412425 -0.15402503 -0.15392594 -0.1538270 -0.1537281
> select(ridge)
modified HKB estimator is 1.028035
modified L-W estimator is 1.051186
smallest value of GCV at 1
> ridge$coef[,ridge$lam == 0.9]
      x1_      x2_      x3_      x4_2      x4_3
0.23841435 0.14948565 -0.09793568 -0.11046006 -0.15392594
```

Lamda parametresine göre katsayılar değişmektedir. Hızlı azalış ve artışların bittiği yerde bir lambda parametresi belirlenerek katsayı kestirimleri elde edilir ve model kurulur. Lamdanın 0,9 olarak alındığı durumun sonuçlarına göre katsayı kestirimleri aşağıdaki gibi elde edilmiştir.

```
> ridge$coef[,ridge$lam == 0.9]
      x1_      x2_      x3_      x4_2      x4_3
0.23841435 0.14948565 -0.09793568 -0.11046006 -0.15392594
> |
```

Hızlı artış ya da azalış gösteren katsayılar karşılık gelen değişkenler çoklu bağlantılı değişkenlerdir. Değişkenler birbirine yakın değerlerdedir. Sıfır eksenin civarında seyreden değişkenler de modelde önemsiz değişkeni göstermektedir. Burada x3_ değişkeni sıfıra çok yakın olduğu için modele katkısı anlamsızdır.



Açık yeşil ve siyah izlere denk gelen değişkenler çoklu bağlantıdan etkilenir. Mavi, lacivert izlerin değişkenleri çoklu bağlantıdan etkilenmez. Kırmızı iz ise önemsiz değişkeni temsil eder.